John Killeen Commemorative Lecture 2004 'In Pursuit of A Culture of Evaluation'

Michael White, Policy Studies Institute

John Killeen made important professional contributions through his teaching, through his influential activities in careers education and adult careers guidance, and through his researches, where (among other work) he carried out pioneering studies in the field of public policy evaluation[1].

My discussion focuses upon this last area, and is intended to take forward John Killeen's contribution in a positive way. More specifically it concerns the idea of 'a culture of evaluation', which he advocated in the early 1990s (see Killeen *et al.*, 1992). The notion of a culture has been widely debated in varied contexts, including for example corporate or institutional cultures or community cultures. Most discussions of the concept agree that a culture involves both beliefs (or values), and practices. So a culture of evaluation is not just a system for putting evaluation into regular practice, essential though that is. It also requires belief in evaluation as an activity with the potential to do something worthwhile, a belief which underpins and motivates practice.

The 'culture of evaluation' concept is therefore important in requiring some specification of the underlying beliefs or values which evaluation is intended to serve. This is itself a counterweight to the prevalent emphasis on technical issues in discourse about evaluation. The first half of my discussion, therefore, considers why evaluation is important, what it can do for us and why we need it. It will also be necessary to consider an apparent barrier to setting up a culture of evaluation: the widespread current scepticism about the benefits of evaluation, which in effect denies any value which may be claimed for this activity.

If evaluation is important, the best way to praise it is to practice it. John Killeen knew that it was hard to practice evaluation. The last sentence of the last published paper that he wrote was

> 'Rather than do poor evaluation, we could prepare to do it well' (Killeen, 2004).

It is the word 'prepare' which is particularly resonant with me. To do well what is difficult requires purposeful, long-term development. We must apply a long-term perspective for achieving the goal of a culture of evaluation. Furthermore, the idea of a culture of evaluation - an essentially *social* idea - tells me that no individual will get there alone and unaided. Good evaluation will involve a cooperative effort.

In the second half of this talk, therefore, I will discuss how this development might be pursued - just what kinds of preparation are needed if evaluation is to be well done? I will argue that evaluation will not work fully unless government, the central civil service administration, the research community, and practitioners all play their parts in ways that interlock and support one another. I will sketch some of the practical requirements that, in my view, are needed to make evaluation into a successful social activity, in the long term.

When John Killeen put forward his idea of a culture of evaluation, he was specifically thinking of careers education and guidance. But I see it as having much wider relevance. This presentation is couched in general terms and relates to public policy evaluation as a whole. I leave it to those who work in the field of careers education and guidance, or indeed in other fields of practice, to judge the relevance of the points made to themselves, and to apply them where they can.

Why does evaluation matter?

Let me start with an evaluation manifesto, made up of four interlocking claims which I will subsequently try to justify.

1 Innovation in public services is desirable.

2 There is a need to establish what difference each innovation makes.

3 In public services, evaluation is the only way to get this knowledge.

4 Evaluation matters because innovation matters.

Note that, in drawing up the manifesto in this way, I am focusing on a particular kind of evaluation. There are many other kinds of evaluation which can also be worthwhile in other ways, but the particular type I am considering is that which establishes the gain from an innovation.

Why does innovation matter, specifically in public services? The pragmatic argument will be all too

---

[1] John Killeen's work is celebrated in *Career Research & Development*, No. 11, Autumn 2004.

familiar. Public services are under intense pressures from a citizenry that demands higher levels of quality but grudges the taxes to fund provision. These contrary pressures cannot be reconciled without innovatory changes in provision and methods. Unhappily, the existence of these pressures is often used by policy makers and policy formers as a kind of threat. Public services are commanded to become more like private services, with the implication that if this does not happen, they will be replaced by private services. These threats are not empty since in many instances public services have been cut or replaced over the past two decades.

But we can surely see better reasons for innovation in public services than compulsion, fear or the urge to survive. Through innovation we can find improved ways of meeting people's needs, and especially the needs of people who, facing difficulties, turn to the public services for help. The desire to improve the well-being of clients is central to the ethos which is shared by public service workers of all kinds and provides a natural motive for innovation. Of course governments or their agencies may want to innovate for other reasons - for instance, to cut exchequer costs or increase economic performance. But aims of that type will never draw together a wide range of public service practitioners in a common culture.

Innovation, even when practised in a positive way rather than through external compulsion, can have a down-side. For instance, it often involves more risk and more stress. We also have to distinguish innovation which has a worthwhile aim from the cynical uses to which it is sometimes put. The point is <u>not</u> to claim that innovation is a self-evident good or an unmixed blessing. In committing to any value, choices and trade-offs have to be made. A group, an institution or a society which chooses innovation for its positive aspects despite awareness of its down-side is on the way to establishing a culture of innovation.

The need for evaluation follows directly from a commitment to innovation within the framework of public service. Indeed, this is one of the ways in which public services profoundly differ from the market sector. Within the market sector, value to the customer is indicated through prices, and this means that business organisations can assess many kinds of proposed change in a relatively simple way, through information about prices, costs and profits.

Instead, public services have to address the value of services to clients in a more direct way, which inevitably varies with the kind of service and the kind of need which it is fulfilling. This is much more difficult, and often messy; certainly it cannot be reduced to a simple formula. None the less we must persist in asking, of any change in service, 'What difference does it make for clients?' Evaluation is the way of answering questions of this sort, and the only way that is consistent with the ethos of increasing the well-being of clients.

Within a culture of evaluation, service innovations would be assessed in a systematic way and it would become possible to acquire learning about the benefits to clients of different types. This would feed back into choices about ways to improve services and practice, and would also stimulate further innovation as the beneficial results of previous changes became more reliably known. So the culture of evaluation implies a virtuous circle between the value of innovation and the practice of evaluation.

Why scepticism about evaluation?

It may seem that, in asserting the worth of innovation allied to evaluation evidence, I am pushing at an open door. The New Labour government, soon after it took office in 1997, emphasized its belief in 'evidence based policy' and, more generally, its desire to implement 'policies that work' rather than policies selected on ideological grounds. Under the governments of 1997 and 2001, much larger sums have been invested in evaluation research than ever before in Britain. Yet it is plain that neither evidence based policy nor evidence based practice has become widely embedded. Instead, I detect in many quarters a growing public and practitioner scepticism about evaluation. Why has this happened?

One problem, I suggest, is confusion. The word 'evaluation' can quite properly be applied to a huge range of activities, from personal judgements to the results of complex social experiments. While there is a great deal of activity labelled evaluation, little of this generates knowledge about what difference innovations make to clients' lives or well-being. Some of it tells us how initiatives are operated or managed, some tells us about acceptability to clients but not about changes in their well-being. Some, regrettably, is what John Killeen referred to as

> 'the 'eyewash' of favourable but superficial description, or the 'whitewash' which obscures faults' (Killeen, 1996a: 331).

Especially, evaluation is confused in the public's mind, and in the minds of many practitioners too, with *performance measurement*. If there is one thing on which the general public and public service practitioners agree, it is that there are too many service targets and too many performance measures. In reality, the public sees only a fraction of the performance measures which public service practitioners now face.

Performance measurement or monitoring does *not* show which changes in policy or practice make a difference to clients. Moreover, performance monitoring systems too easily become sticks for beating, or carrots for rewarding - often entangled with management and staff appraisals, or with battles between local and central managements. One-off reviews, of the kinds often carried out by business consultants, are still more political and threatening. It is this type of situation, and its bracketing with evaluation, which I believe John Killeen had in mind when he wrote

'Evaluation is tangled up in the macro-politics of national resource allocation and the micro-politics of organisational preferment' (Killeen, 1996a: 331-2).

Even when innovation takes place, and evaluation thoroughly assesses the benefits for clients, there often appears to be a lack of connection with 'evidence based policy'. When New Labour came to power in 1997, its flagship programme was New Deal for Young People (NDYP) and the large evaluation effort devoted to this programme was trumpeted to announce the advent of evidence based policy. NDYP was indeed a highly innovative programme which was a worthy focus of evaluation effort. But NDYP was immediately launched as a national programme before any evaluation had even commenced. By the time findings from the evaluation came through, it was too late to alter the design of the programme substantially in the light of its more and less successful features. One or two years devoted to piloting the major features of the programme in various combinations would have provided a wealth of evidence on which future policy towards young unemployed people could have been based.

It can of course be argued that the new government was justified in launching immediately into a large initiative which, quite apart from its direct effects on clients, would set an innovative tone for its time in office. However, the story of NDYP is not an isolated one. Other labour market and educational programmes have been introduced in pilot form but greatly extended or rolled out into national programmes before the evaluation evidence was in, and in some cases contrary to the evidence itself. For instance, one could point to the decision to introduce a national programme of work-focused interviews for lone parent benefit claimants, before completion of the pilot evaluations and despite rather negative results from those evaluations[2]. In the face of events such as these, scepticism or even cynicism about evidence-based policy is natural.

But if scepticism is understandable, it is not altogether just. An agenda of evidence-based policy which is only haltingly underway is a distinct improvement over no such agenda. The lack of policy influence from evaluations such as NDYP has been disappointing, yet these evaluations may provide useful practical lessons that transfer to other programmes in the longer run. There does exist commitment to innovation in the public services and to evaluation, and there is the chance of linking those commitments together in a more effective way. In the second half of the talk, I discuss steps to embed those commitments more firmly in practice.

Steps to a culture of evaluation

To develop a culture of evaluation in practice involves, as I envisage it, four institutional groupings. These are:

* the government in power
* the civil service departments and their agencies
* the research community
* the practitioners who organise and provide services to the public.

What each of these institutional groups does about evaluation affects what the others can do. I will discuss each group separately and in turn, but throughout I am assuming that the eventual aim is a *cooperative* type of development. Broadly speaking, I perceive a current situation where the first three groupings are already engaging with evaluation of public service innovations, albeit imperfectly, but the fourth group - practitioners - remains largely on the outside, or takes on a very small role[3]. Accordingly, the sub-plot of what I have to say is to increase practitioner participation. This step turns out to have great implications for the whole evaluation agenda.

The government in power

The government (by which I mean the political party in power) is and will be in the foreseeable future committed to innovation in public services. Since the early 1980s, bold initiatives have been seen as the way to make an impression on the electorate, and it is hard to envisage any turning away from this.

But it is wholly understandable if impartiality and self-criticism are hard for government. Government is and must be political. It has a natural tendency to inflate evaluation results which declare its initiatives effective, while 'overlooking' evaluation results which say the opposite. Yet a government role is essential in developing a culture of evaluation, not just to secure funding, but also because research which does not engage with government has a diminished influence on policy. The need is to maintain government's interest in evaluation but, through the roles of the other institutional groups, to shape that role to be more accepting towards critical evidence.

It is worth reflecting, for a moment, on how New Labour got to promote the idea of evidence-based policy. Their underlying desire, surely, was to find a way of justifying policies without being trapped in conventional left-of-centre political stances. A 'doing what works' stance protects against accusations of ideological bias to which Labour had long been vulnerable. So evidence based policy offers potentially greater freedom and security of policy choice. But this in turn opens up a debate about effectiveness and delivery, in which all three main

---

2 Work-focused interviews for lone parent claimants were introduced nationally from April 2001 onwards. For the evaluation of the pilot program, see Kirby and Riley, 2003.
3 The main exception is in health services.

political parties are now embroiled. It is in engaging with this current discourse about effectiveness that the other social actors can influence government towards a greater commitment to evidence.

Here the media could play a key role, but only if they became more knowledgeable. Manifestly, media reaction has a large influence on the calculations of the government in power and the parties in opposition. For instance, all political parties have become more circumspect about making promises which cannot be funded from current or foreseeable tax revenues, because media commentators have become alert and critical on this point. The media have undoubtedly been 'educated' on this issue by analysis provided by the research community, notably in this instance the Institute for Fiscal Studies.

In contrast, on claims about the effectiveness of government initiatives or programs, the media frequently remain gullible. As recently as September 2004, The Independent reported as a government achievement that 500,000 jobs had been entered by participants in the New Deal for Young People. Yet the evaluation evidence which the government itself commissioned indicates that no more than 1 in 10 of these jobs is additional and therefore attributable to NDYP: the remainder would have happened anyway (see White and Riley, 2002).

A better informed media would challenge the government when it disregarded its own evaluation results. Another incentive for government to wait for the evidence before committing to a policy is that this would render the policy less vulnerable to being reversed by a subsequent government. Evidence-based policy takes longer to create, but should represent a more lasting achievement of the government which brings it to pass.

How might government behaviour change as it became drawn into a culture of evaluation? One possible model is the USA, where economic evaluation has been developed strongly since the 1970s. Most US programmes now start as 'demonstration projects' which are managed, in conjunction with local public service providers, by external contractors who are also responsible for providing or arranging evaluation. Thus the delivery and evaluation of these pilot programmes are at more than one remove from central government. Only when the evaluation results are in - a process which often requires several years - will scaling-up be considered. So government adopts a wait-and-see approach in which it must be *persuaded* of benefits before it commits large-scale national programme funding. In so doing, of course, it also protects itself from being blamed for failing programs. It is notable that the first demonstration projects that follow the US pattern have recently been launched in Britain - most notably the Employment Retention and Advancement Demonstration (ERAD) project. How this project is treated by government will be an acid test of progress towards evidence-base policy.

It will be particularly interesting to see whether government has the patience to complete the project before deciding its policy.

Does this type of development, with the government in power becoming more cautious about its own innovative commitments, imply less innovation? Not necessarily. As government steps back a pace or two, more space is made for other social institutions to innovate more. Contacts that I have had over the years with local agencies and voluntary bodies in the USA (and also some other countries, such as Australia or the Netherlands) have impressed me by the vigour and enthusiasm of their innovative work. The culture that has developed in the USA is one where the centre learns from local experimentation rather than imposing its will across the board. Of course, this reflects a decentralised governance tradition which is different from that of Britain. We cannot copy the US system, but we can draw confidence from it about the capacity for innovation which exists throughout society.

The civil service

Civil service departments and agencies have played a major role in fostering public service innovation and evaluation. For example, in the 1980s the Manpower Services Commission and its successors played an important part in reforming adult skills, with the Employment Service playing a similar role in relation to services for unemployed people (the story is told in Price, 2000). Over the past two decades, HM Treasury has been the chief moving force in establishing the need to evaluate new initiatives and in demanding higher standards of evidence. Departments and agencies long exposed to evaluation have also come to see that an evidence-based culture strengthens their position, since they are responsible for marshalling the evidence.

This background means that the civil service will be, for the foreseeable future, key players in any culture of evaluation which one may hope to establish. But those central departments and agencies which have been leading the way in evaluation need to think more about how they can foster a culture of evaluation which goes beyond their own walls.

One way in which they can do so is in developing what might be called an 'information infrastructure' for evaluation. How can one properly evaluate a service innovation, if one does not record what service is actually delivered to the client? Again, evaluation can be greatly improved if details about the clients, and what happens to them subsequently, can be extracted from administrative records that are created along with the service, rather than obtained through costly (and frequently flawed) special surveys. Such a development democratises evaluation by making it less costly and therefore possible for more people to get involved with and use. Here the news is

generally encouraging, with many steps currently being taken in this direction.

The central administration should also be giving thought to developing a wider capacity for both innovation and evaluation. It is an unfortunate part of the British governance tradition to rely heavily on the central elites. Much of what I want to say about decentralisation will come in the following sections, when I discuss the research community and the practitioner community. But it would help if the central administration wanted to develop the role of these groups and give them space to be independent actors. Again, there are some encouraging signs. In the ERAD project, which was referred to in the previous section, front-line staff serving unemployed and in-work benefit claimants are being trained to play an active role in the experiment, and similar developments are visible in some other demonstration projects and in some community initiatives. Some departments are encouraging practitioners to set up their own D-I-Y evaluation activities and providing advice on how to get started. This at least shows awareness of a need for practitioner involvement.

The research community

In developing a culture of evaluation, the research community has two main roles. One is to establish sound standards and methods of evaluation, so that it can be trusted, while the other is to make evaluation and its results accessible to a wider community of interests, and especially the practitioner community.

While the amount of evaluation which is technically of a high standard is growing, this tends to be concentrated in a narrow circle of specialist research institutions who chiefly talk to each other and to their specialist opposite numbers in the central administration. Evaluation is at risk of developing as a divisive rather than an inclusive institution. Thus, the emergent issues for the research community concern collaboration and interaction with practitioner and research-user groups.

As an example of what might be achieved, John Killeen pointed to the activities of applied psychologists in the USA, who are heavily involved in guidance and counselling practice, often devise new methods to be tried in those fields, and also lead the way in evaluating the effects of those innovations (Killeen, 1996b; 2004). In Britain, however, researchers involved in evaluation tend to see innovation and practical application as the responsibility of the government departments, so that they do not need to interact with practitioners. But it is only through researcher-user interaction that higher standards of evaluation will become accepted and understood in practitioner groups.

An encouraging sign of the development of such interaction is the establishment, in 2004, of the National Guidance Research Forum (NGRF), a joint initiative between the Guidance Council, the National Institute for Careers Education and Counselling, the Institute for Employment Research at the University of Warwick, and the Centre for Guidance Studies at the University of Derby. The Department for Education and Skills is providing initial funding, which is another indication of growing awareness at the centre of the value of fostering practitioner involvement in research.

There is also a need for the research community to play a part in developing and disseminating evaluation methods which are relatively simple and accessible, yet sound. This would contribute to the aim of practitioners developing their own capacity to evaluate innovations. This is a subject which is too large for me to cover here (for instance, it inevitably leads into the controversial issue of social experiments). But it is worth pointing out that it was with the aim of encouraging practitioner involvement that John Killeen advocated the development of measures of the 'learning outcomes' of guidance. More discussion of this follows in the next section.

Practitioner groups

The purpose of evaluation is to find which innovations make a worthwhile difference, but implicitly we only make this our aim because we want to put those innovations *into practice*. This cannot be achieved without the involvement of those who deliver the services.

In the formulation of new evaluation projects, and in appraising the results through feedback, there is the potential for more practitioner involvement, and this already takes place in some cases. But, from the nature of things, this will nearly always be in the form of practitioner representatives and practitioner experts, rather than anything more widespread. For a culture of evaluation to take hold in a public service, the potential evaluation roles of *all* its practitioners need to be developed.

One of the most important practitioner roles exists even before any evaluation takes place. Practitioners play the key roles in information systems on which evaluation draws, and the potential of this information role has been greatly underestimated. Indeed, lack of appreciation of this information role is one of the chief reasons why a culture of evaluation has not yet developed in most public services.

This can be illustrated with an anecdote. When John Killeen and I designed an evaluation of adult guidance services (see Killeen and White, 2000), we assumed that clients would be recruited into the project by advisors when they first attended, and that the advisors would arrange for these clients to complete an initial questionnaire about their pre-guidance circumstances. However, during piloting it rapidly became apparent that the procedure was impractical. Advisors lacked the knowledge to adapt their normal reception procedure in

this way, and local managers lacked the time or knowledge to train them to do it. Since we ourselves did not have the time or funding to devise training, we had to revert to a research method which removed responsibility and involvement from the front-line staff.

This example is <u>not</u> meant to blame front-line staff or their management in any way. Rather, the fault lay with our own mistaken assumptions, and in a lack of prior training without which staff could not play their part in research procedures. Then researchers devise ways of gathering their data independently, so that the front-line practitioners become excluded from the evaluation process. If this happens in the case of skilled professional staff, as in careers guidance, it is still more likely where the front-line staff have relatively low levels of professional training, as in the case of (say) Jobcentre advisors. But there are some encouraging signs of change being fostered from the centre. In the ERAD project which I referred to earlier, much attention is being given to providing the training which will enable front-line staff to induct clients into the pilots and into the evaluation process.

A related but more complex issue is the embedding of tests or other standardised measures into the practitioner relationship with clients. Anyone who consults a doctor or dentist expects systematic observations to be made and recorded, but elsewhere this often appears to cause difficulty. An example came in an evaluation of innovative provision for adult basic skills courses, another project where John Killeen and I worked together. Part of the evaluation design, as specified by the DfES, was that local educational providers (mostly FE colleges) would carry out pre-course testing, with standard national tests, to establish baseline levels of attainment. This was also to be done in comparable courses which lacked the innovative features. In this way, gains in attainment from the innovative courses could be compared with those from traditional courses. In practice, however, this part of the evaluation had to be dropped since large proportions of learners were not given the pre-tests by their tutors.

This issue of embedding standardised measures is crucial for John Killeen's hopes of using 'learning outcomes' as a way of evaluating innovations in provision (see, especially, Killeen, 1996b). Although he put this forward specifically for careers guidance and counselling, the idea is more general. The example I have just given about adult education obviously concerns learning outcomes, and many other kinds of service - for instance, employment offices, pre-natal advice, or probationary supervision - have learning functions. Once practitioners routinely collect measures of learning outcomes, or similar outcomes that come directly from the service provided, evaluation would become easier and less costly to mount. Equally important, practitioner groups could become more closely involved in the process of evaluation itself, rather than (as at present) usually being excluded from the loop.

However, developing useful and reliable measures of learning or similar outcomes is a far from easy task. Even where learning measures exist, they will not be consistently put into practice if staff lack confidence in them or if they are not user-friendly. If such measures are to become embedded in practice, therefore, involvement of practitioner groups *in the development stages* of the measures themselves needs to be achieved. How tests or other measures are applied in the client setting is an important issue and one where practitioner involvement and feedback should improve matters greatly.

My emphasis on the practitioner roles in information and measurement might seem to confine them to a relatively routine or humdrum position in the culture of evaluation. But this is not so, because the information and measurement roles are only the beginning. One of the results of a public service having embedded measurement is more scope for local or small-scale research, offering more opportunity for practitioner involvement. Consider the health services, which perhaps offer the most developed example of a culture of evaluation. Medical, nursing and paramedical staff learn about research procedures as part of their professional training, and standard measurement and recording procedures are used as a matter of routine. This leads to a situation where a great number of small-scale clinical trials are in progress all the time, and these are widely dispersed around the health services.

It is true that evaluation remains a difficult task, and practitioners cannot grow into a full research role overnight. But if evaluation is seen as a collaborative activity, then there are various ways in which practitioners might combine with researchers so as to develop capacity. There can be researcher-led projects with practitioners as consultants, there can be joint evaluation teams, or practitioners can plan and manage their own evaluations while bringing in technical services from researchers. Hopefully, the central authorities could be persuaded to push more of the funding outward to support such collaborative developments.

Finally, it is important to consider the implications of all this for practitioner involvement in innovation. The capacity to evaluate ideas, and so establish their effectiveness, surely is an incentive to be interested in innovation and, in short, to innovate. A dispersed capacity to evaluate should therefore lead to more widespread innovation with practitioner involvement. This would also tend to produce a more balanced kind of innovation. Alongside the large central initiatives, of which there are perhaps more than enough, there would be a growth of smaller innovations, concerned not so much with the grand structure of service provision, but with the detailed 'how?' of service delivery. In the long run, it may be that these detailed improvements in the methods of assisting the clients of public services will make more difference

than the great changes in the organisation of provision that have been so conspicuous in the past 20 years.

An increased role for public service practitioners in both innovation and evaluation cannot however be fully realised without professional training. Bearing in mind Alexander Pope's famous dictum about fools and angels, I will not prescribe how training should be modified for any practitioner group, let alone for all of them. However, I would suggest to those responsible for practitioner training that they might consider how greater involvement in innovation and evaluation - the R&D of public services – could be fostered within training provision.

Conclusion

In this discussion I have tried to take forward the idea of a culture of evaluation which was originally advocated by John Killeen. In brief, I see a culture as consisting of both commitments and practices, understood and shared between social groups. I then apply this concept to evaluation in the public services.

The first half of the discussion argued that the main point of evaluation is to identify innovations which make a positive difference for clients of the public services. A culture of evaluation would draw its energy from wide agreement that improved services are worth pursuing, and that it is therefore worth finding out *what* makes a difference for clients.

The second half of the discussion concerned the practice of evaluation across the chief institutions concerned: the government in power, the civil service departments and agencies, the research community, and practitioner groups in public services. I have focused on changes in attitudes and behaviour which would help to form a culture of evaluation which links all these groups, and I have emphasised the importance of involving practitioners if a widespread culture of evaluation is to become a reality. What makes it difficult to establish a culture of evaluation is that no one group can accomplish the task. It requires each group to take positive steps while being aware of and respecting the parts played by the others. It is through an accumulation of such positive steps that a culture of evaluation can be built, and as stressed at the outset of the discussion, this is sure to require long-term persistence.

John Killeen's original idea of a culture of evaluation is, I believe, an important one and, as I have tried to develop and apply it, I have come increasingly to see the seriousness of its implications. Rather probably my proposals have not measured up sufficiently to the task, and they are by no means the only ones which might be suggested. I hope that others may take forward the debate in pursuit of a culture of evaluation.

References

Killeen, J. (1996a). Evaluation in Watts, A.G., Law, B., Killeen, J., Kidd, J.M. and Hawthorn, R. (eds.) *Rethinking Careers Education and Guidance*, London: Routledge.

Killeen, J. (1996b). The learning and economic outcomes of guidance in Watts, A.G., Law, B., Killeen, J., Kidd, J.M. and Hawthorn, R. (eds.) *Rethinking Careers Education and Guidance*, London: Routledge.

Killeen, J. (2004). Evaluation: Principles and Starting Points, *Career Research and Development*, No. 11, 3-12.

Killeen, J. and White, M. (2000). *The Impact of Careers Guidance on Adult Employed People*. Research Report RR226, Sheffield: Department for Education and Employment.

Killeen, J., White, M. and Watts, A.G. (1992). *The Economic Value of Careers Guidance*, London: Policy Studies Institute.

Kirby, S. and Riley, R. (2003). *The employment effects of full participation in ONE*, Part Three of *The Final Effects of ONE*, Research Report RR183, London: Department for Work and Pensions.

Price, D. (2000). *Office of Hope*, London: Policy Studies Institute.

White, M. and Riley, R., (2002). *Findings from the Macro evaluation of the New Deal for Young People*, Research Report RR168, London: Department for Work and Pensions.